





Moving along the spectrum of {Data -> Information -> Knowledge -> Wisdom} requires monotonically increasing levels of human attention. Attention is required to cross the interface between information and knowledge. Taking the final steps from knowledge to wisdom remains a uniquely human endeavor.


The traditional, and still vexing, challenge is getting critical information to those who need it in a sufficiently timely fashion that it can contribute to the quality of the decisions they make. This problem is made more complex given the accelerating rate of scientific and technical discovery, typified by the ever shortening time period for the doubling of information (currently estimated at 18 months).

The Information Management program seeks long-term fundamental breakthroughs in our ability to acquire, organize, utilize, and preserve valuable information.




Information Management Current Environment






USA Today - July 2, 1996

“History is full of cases where things could have been avoided, but the right person never found out until it was too late.”
- Dataquest analyst Christine Ferrusi Ross





Copyright 1996, USA Today. Reprinted with permission.

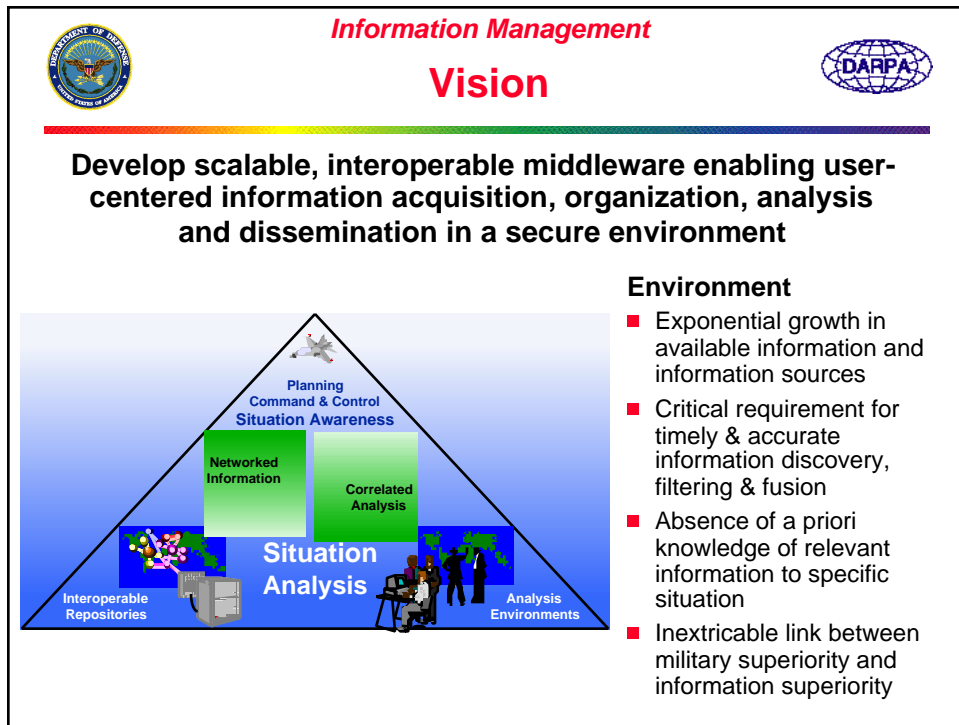
“Months before ValuJet Flight 592 crashed, the FAA had the safety data it later used to ground the airline.”

“In Saudi Arabia, US authorities wanted to widen the security zone around the Khobar Towers housing US troops, but the Saudis turned down the request. Defense Secretary William Perry didn’t learn about the denial until 4 days after a terrorist bombing killed 19 Americans.”

“A sharp rise in arson incidents at black churches in the South was not detected until February, partly because the system for collecting fire data is haphazard.”

In each case, critically important information was collected but not organized or analyzed in a way to give it meaning.

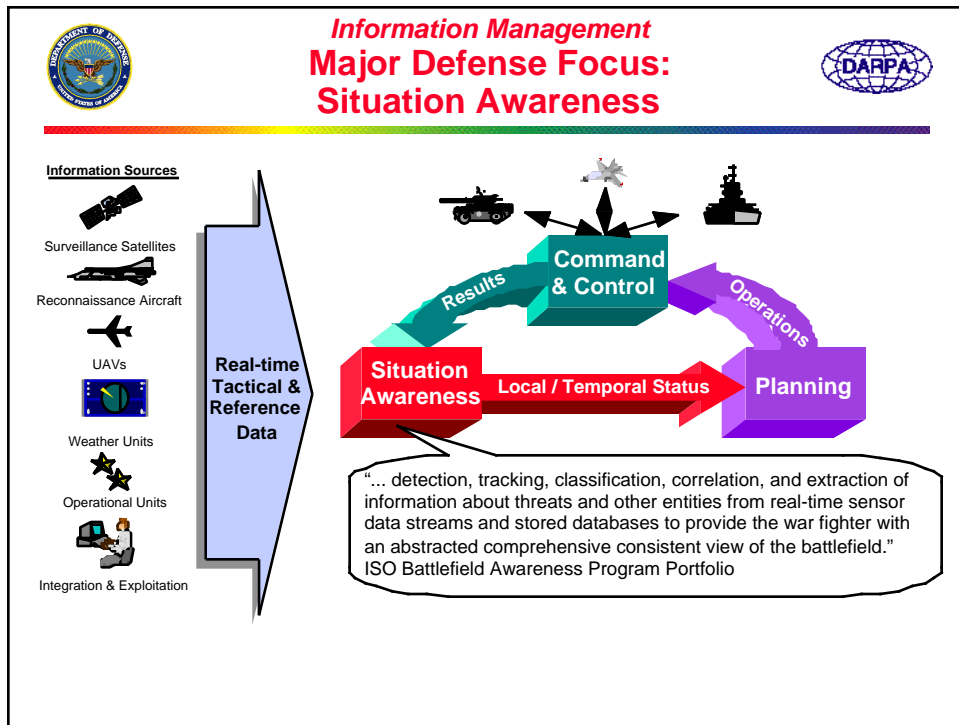
Each of these cases addresses failures in vertically integrated systems... it cannot be guaranteed that critical information will even exist along the vertical path; all sources of information need to be available and usable... competing for appropriate attention among decision makers. As decision/action rates increase, the need for timely and comprehensive information increases, but *timely* and *comprehensive* are competing goals.



The Information Management (IM) program proposes major advances in acquiring and effectively using vertically-integrated, as well as horizontally-distributed information resources to provide the defense analyst with a comprehensive ability to assess a rapidly changing situation.

Its products will be scalable, interoperable middleware:

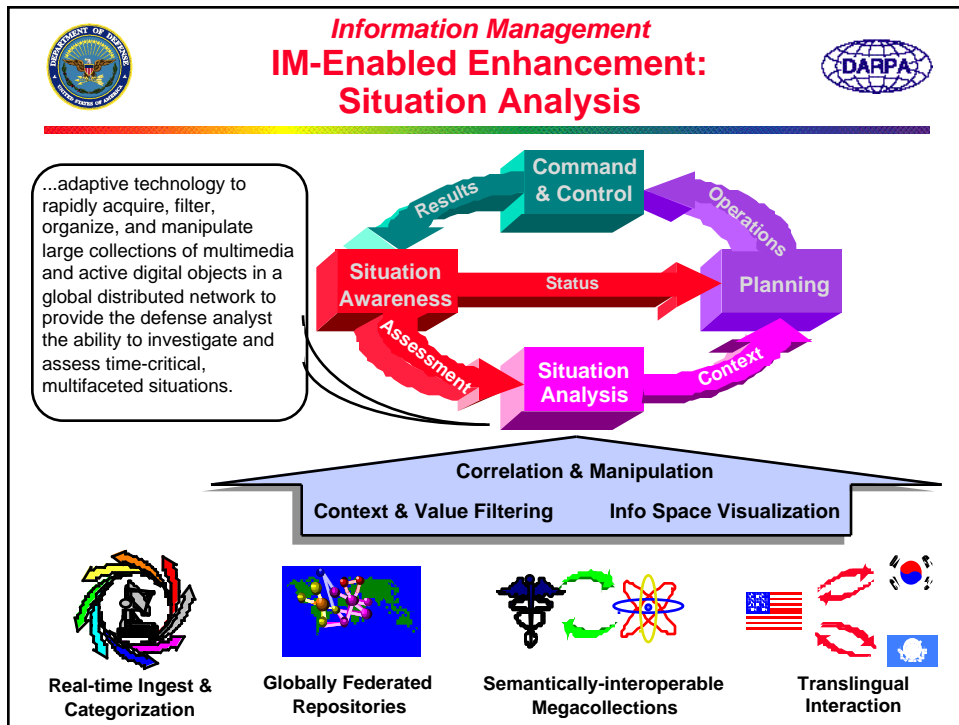
- to manage exponentially growing information resources
- to focus the analyst's attention on highly relevant materials
- to organize information for rapid exploitation in unpredictable circumstances
- to provide superior ability to evaluate all aspects of a given situation to appropriately inform rapid decision processes



Situation Awareness, or *Battlefield Awareness*, including "detection, tracking, classification, correlation, and extraction of information about threats and other entities from real-time sensor data streams and stored databases to provide the war fighter with an abstracted comprehensive consistent view of the battlefield," is a key requirement for the Defense community. Precise and timely situation awareness leads to appropriately targeted plans, which can subsequently provide the base for command & control, leading to successful accomplishment of objectives.

Note that the key objectives of situation awareness are:

- vertical integration of multiple sources of information
- extraction of relevant information
- abstraction of extracted information into semantic units
- timeliness with respect to decision processes
- consistency across diverse information sources



The IM Program enhances situation awareness with horizontal capabilities of correlation and manipulation across disciplines and among languages:

- Filtering information by relevancy to the problem context and value of content to the user
- Visualizing complex and abstract information spaces for highly facile navigation and retrieval

The program seeks to develop adaptive technology to rapidly acquire, filter, organize, and manipulate large collections of multimedia and active digital objects in a global distributed network to provide the defense analyst the ability to investigate and assess time-critical, multifaceted situations.



Information Management

Goal



Increase the defense analyst's ability to aggregate & evaluate all available networked information resources to investigate & assess time-critical, multifaceted situations

- Establish capability for real-time situation analysis
- Demonstrate adaptive technology for rapidly acquiring, filtering, organizing, and manipulating very large multimedia and active digital objects in a global distributed network

NCSTRL (~40)

"Research Library"

Annotated, complex multimedia

Capability	Present	Goal
Federated Repositories	Tens (Custom)	Thousands (Generic)
Items / Repository	Thousands	Millions
Size of "Large" item	1 MB	100 MB
Typical response time	10 seconds	100 milliseconds
Mode	Play & Display	Correlate & Manipulate
Interoperability	Syntactic	Semantic
Filters	Bibliographic	Contextual
Language	Multilingual	Translingual
Content extraction	Forms & Tags	Semi-structured

DLI testbeds

Book

WWW Retrieval

Dependent on structural characteristics, textual similarity, histogram matching, etc.

e.g., layered, multivalent documents

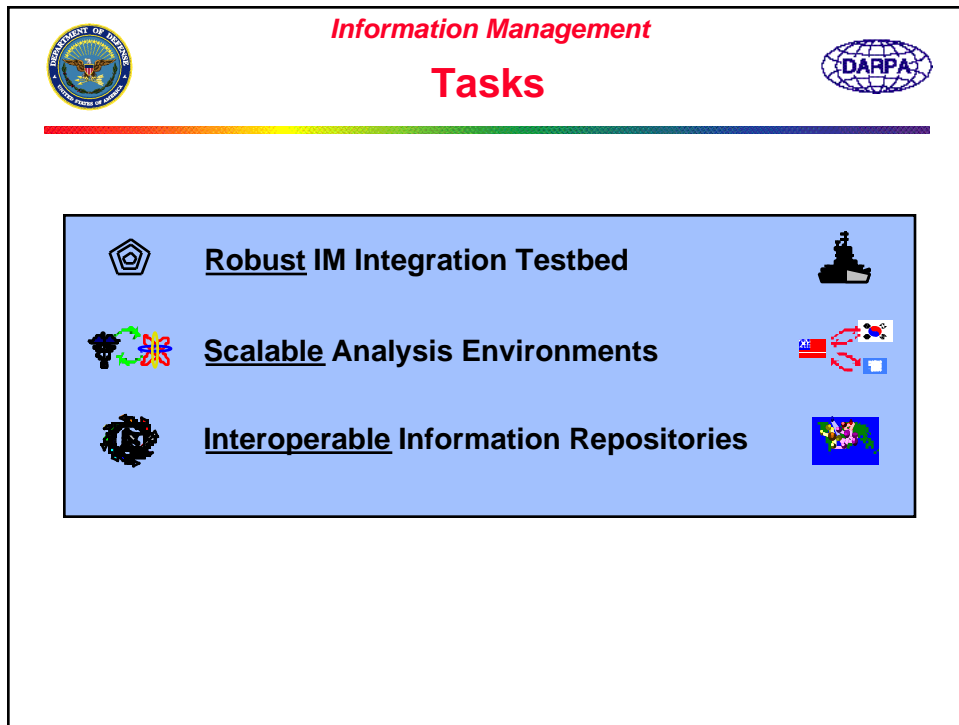
Includes content & concept-based access

Title, author, subject keyword, date, authority, ...

Task-relevance, value-based

The program's goals are aggressive, seeking typically two orders of magnitude improvement in key parameters: numbers of coordinated repositories, sizes of collections, sizes of objects, and timeliness of response.

In addition, the program seeks qualitative improvement in the ability to create, correlate, and manipulate information from multiple disciplines and in multiple languages.



The IM Program includes three tasks:

- The IM Integration Testbed is designed to bridge between fundamental, focused research and the need for evaluation in Defense-related applications.
- Analysis Environments provide the end-user facilities for correlation and manipulation of diverse information resources.
- Information Repositories provide the secure and interoperable facilities for information organization and management.



Information Management



Task 1: Robust IM Integration Testbed

Integrate and evaluate repository and analysis environment technologies in Defense scenarios

Challenge	Approach
<ul style="list-style-type: none">Evaluating diverse technologies developed by distinct research teams in the context of realistic Defense-relevant scenarios	<ul style="list-style-type: none">Establish Information Management (IM) Integration Testbed to assemble and assess the feasibility of emerging capabilities with appropriate Defense partner(s)

In the long path from laboratory to operational deployment, innovative ideas can easily get lost for lack of an appropriate transition mechanism. While it is ineffective to clutter a research agenda with too many operational details, it is likewise ineffective to introduce a research product to the ultimate user community prematurely, even if this occurs through a traditional, large-scale DoD evaluation facility.

The IM program bridges the gap between research program and DoD evaluation facility through a separate, modestly sized integration testbed. The IM Integration Testbed will work closely with IM researchers to identify promising technologies for integration into an experimental IM environment, in which the feasibility of including the new technologies in Defense applications will be tested.

Early results will be fed back to the research community to guide further research. As capabilities mature in the testbed environment, arrangements with DoD evaluation facilities will be made to test the new technologies in more realistic Defense scenarios, situations, and facilities.



Task 2: Scalable Analysis Environments

**Develop active information analysis environments with
translingual semantic functionality**

Challenge	Approach
<ul style="list-style-type: none">• Acquiring information resources based on semantic (e.g., conceptual) rather than syntactic (e.g., textual) content• Visualizing a complex information space for effective navigation• Limiting the breadth of selectable resources by considering the intended use or user• Providing capability to manipulate multimedia & complex documents across a language barrier	<ul style="list-style-type: none">• Extend statistical algorithms to support vocabulary switching & object categorization across repositories and collections• Include concept spaces and semantic category maps in visualization metaphors• Develop value-filtering techniques and integrate these into a spectrum of pre & post-filtering tools• Integrate advances in machine translation technology with authoring & analysis tools for layered document structures

Today's information retrieval systems rely largely on indexing the text of documents. While this can be quite effective in bounded domains in which the usage and definition of words is shared, performance suffers when materials from multiple disciplines are represented in the same collection. Rather than being the exception, however, this is typically the rule. We need systems that can look beyond the words to the meaning and the concepts being expressed. Several approaches are promising; some recent success has been achieved, for example, using statistical approaches on large corpora. These and other approaches will be explored and extended.

The navigational metaphor has become ubiquitous for information seeking in the network environment, but highly effective and facile tools for visualizing and navigating these complex information spaces remain to be discovered. Incorporation of concept space and semantic category maps into visualization tools may be a promising improvement, but other approaches, including questioning the basic navigational metaphor are, likewise, of interest.

Query languages and tools seek to identify materials in a given collection which are similar to the characteristics expressed in a given query. But these characteristics focus on the information artifact and have yet to consider non-bibliographic attributes which might serve to focus a search more tightly, such as what kind of individuals have been reading this material, and was it useful to them?

Substantial success has been made recently in foreign language translation within carefully bounded domains. The IM program seeks to develop this capability in the context of networked information, focusing initially on translingual bibliographic functionality.



Task 3: Interoperable Repositories

Develop technology for scalable, interoperable, and secure information repositories

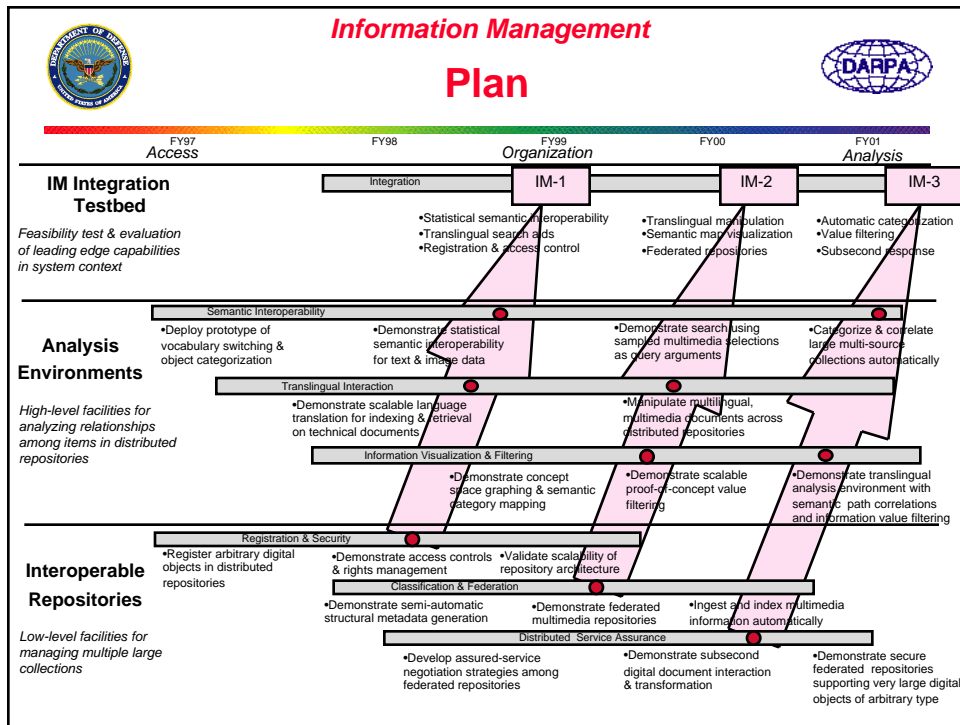
Challenge	Approach
<ul style="list-style-type: none">• Organizing megacollections spanning thousands of federated repositories efficiently for highly expressive access• Extending functional capabilities beyond passive objects (e.g., documents) to include active objects (e.g., services)• Securing sensitive information in open repositories from unauthorized access and utilization• Providing real-time interactive response for very large digital objects	<ul style="list-style-type: none">• Build on DLI projects & testbeds to design scalable distributed repository architecture, implement in generic middleware, and test in post-DLI Integration Testbed• Define indexing and cataloging requirements for active objects, and develop techniques for acquiring metadata• Include requirements for access control and intellectual property rights management in repository architecture• Collaborate with PSOS for interoperability experiments over ITO broadband network

Scalability is a major objective of the research agenda. The challenge is to build scalable repository technology that supports the federation of thousands of repositories, presenting to the user a coherent collection consisting of millions of related items, and to do this rigorously across many disciplines.

The nature of intellectual materials in an electronic environment is expanding rapidly, including multimedia, complex documents, and, more recently, active objects. Catalogs, indexes, and related information management tools are going to need to deal effectively with a very wide and diverse set of materials.

One of the most important and difficult challenges is to provide sufficient protection of individual and intellectual property rights such that information resources with significant monetary, proprietary, or related value-driven attributes can be delivered over an open network environment. These requirements are to be a significant component of the repository research in Information Management.

As the size and complexity of information objects increases, so also does the bandwidth required to utilize these objects. Real-time interactivity is required for the time-critical assessment of complex situations, pushing the bandwidth requirements yet higher. As this capability emerges, broadband interoperability becomes feasible, in which the user's inputs are no longer constrained to a very narrow bandwidth, with the return channel carrying the high volume materials. The IM program will explore the nature of such broadband interoperability and the opportunities it brings to raise the effectiveness of the information user.



This chart shows the overall structure and phasing of the IM program, including how progress in the areas of Analysis Environments and Interoperable Repositories is brought together for test and evaluation in the Integration Testbed.

The work in Analysis Environments includes three major tracks:

- Semantic interoperability strives to extend the analyst's ability to interact with diverse information from distributed sources at the conceptual level.
- Translingual interaction builds on recent successes in machine translation to provide the information user the facility for identifying and evaluating the relevance and value of foreign language materials to a particular query, without assuming the user has any proficiency in the foreign language.
- Information visualization and filtering focuses on the development of improved tools for visualizing and navigating complex multidimensional information spaces, and on user-customizable, value-oriented filters to rank information consistent with the context of the task being performed.

The work in Interoperable Repositories also consists of three tracks:

- Registration and security provides the registration, access controls, and rights management facilities required to support defense-related applications in an open network environment.
- Classification and federation advances the capability to automate the acquisition, classification, and indexing of information resources among distributed repositories.
- Distributed service assurance addresses the vital concerns of matching user interaction styles and needs to system performance capabilities. This work also pushes the boundaries of interactivity over broadband networks.



Information Management



Milestones

Integrate and evaluate information management technology in Defense-relevant scenarios

FY99 Demonstrate Integration Testbed by evaluating emerging analysis capabilities in context of distributed repositories with access controls

FY00 Demonstrate federated repository functionality and evaluate scalability and interoperability of repository designs and analysis environments

FY01 Deploy scalable prototype in Defense application with cross-repository information analysis (semantic retrieval, indexing, value-filtering, interoperability, categorizing)

Develop scalable information analysis environments supporting semantic interoperability

FY97 Deploy net-accessible prototype demonstrating vocabulary switching & object categorization; deploy operational prototype with active references to technical literature to licensed institutions

Investigate application of machine translation to translingual IR in scientific & technical disciplines

Develop information value framework to characterize prior use of objects

Validate electronic copyright management technology in Library of Congress Copyright Office

FY98 Develop algorithms to effectively search collections of documents for words used only in restricted senses

Demonstrate translingual search aids for military or technical documents in English, Korean and a European language

Design query and preferences languages incorporating similarity and value filtering

FY99 Demonstrate semi-automatic topic assignment for unrestricted documents with acceptable accuracy

Develop framework for combined text, image, relational interoperation

Demonstrate translingual query by entering English language query & retrieving documents in at least 2 foreign languages with usable translation of bibliographic information

Demonstrate post-filtering of high recall information discovery algorithms to improve precision to best of integrated IR algorithms

FY00 Demonstrate query-by-example using extracted subsets of multimedia (e.g., image) as search arguments to retrieve multimedia materials

Demonstrate capability to manipulate (e.g., cut & paste) suitably constrained documents across a language barrier

Demonstrate distributed prototype of information-value-based retrieval, where information is generated, stored and indexed in federated repositories

Demonstrate real-time (<1 sec) transformation of digital documents retrieved from large digital repositories

FY01 Demonstrate automatic correlation through categorization of at least one million items distributed among at least 1000 federated repositories

Combine use of image and natural language processing to find information objects across multiple languages relevant to a given natural language query

Milestones have been identified for each of the three task areas, as shown on this slide. These are intended to define target laboratory capabilities which we expect to be achievable within the context of this program.



Information Management Milestones (cont.)



Develop technology for scalable, secure & interoperable information repositories

FY97 Develop testbed for electronic deposit, registration and recordation of digital objects

Demonstrate agent architecture for cross-collection search and results fusion

FY98 Demonstrate electronic document management with access controls, rights management, and document validation in accordance with LC production standards

Demonstrate statistical co-occurrence techniques for texture classification of images

Demonstrate semi-automatic generation of metadata derived from structural attributes of information (e.g., SGML & FGDC tags)

FY99 Validate design of secure repository architecture for digital objects up to 100 MB in size

Demonstrate negotiation among interoperable repositories to set assured performance parameters for access, considering intellectual property terms & conditions

FY00 Demonstrate automated ingest and indexing of real-time multimedia information objects (e.g., video)

Demonstrate sub-second document transactions among federated repositories interconnected by broadband networks

Demonstrate distributed filtering & versioning for archiving digital objects

FY01 Demonstrate security infrastructure for value-based retrieval, in which malicious users cannot spoof trustworthiness of information.

Deploy document recognition services to automatically render structural layers from legacy (i.e., scanned image) documents.

Demonstrate secure distributed repository architecture supporting digital objects (up to 100 MB) of arbitrary type



Information Management



Technology Transition Plan

Research Area

IM Integration Testbed

Analysis Environments

Interoperable Repositories

Transition Mechanism

Evaluate in the context of realistic large-scale systems

Establish integration testbed available to IM PI's and Defense users; Coordinate with DARPA programs (e.g., SUO Situation Analysis & Tasking); Collaborate with Service labs and Services (e.g., NReD & 3rd Fleet Sea-based Battle Lab)

Provide network-accessible services & downloadable software

Demonstrate capability of analysis environments through interoperable DL, NCSTRL, LC & DTIC pilot implementations; Leverage NSF & NASA funding; Leverage DOD investment in language understanding & machine translation; Demonstrate & test in IM Integration Testbed for transition to Battle Lab.

Disseminate repository designs & reference implementations

Deploy for interoperability test & community visibility in multiple heterogeneous domains, including NCSTRL testbed, LC pilot implementation, DL projects, & DTIC; Collaborate with NARA & CNI; Leverage PSOS & Quorum

Legend:

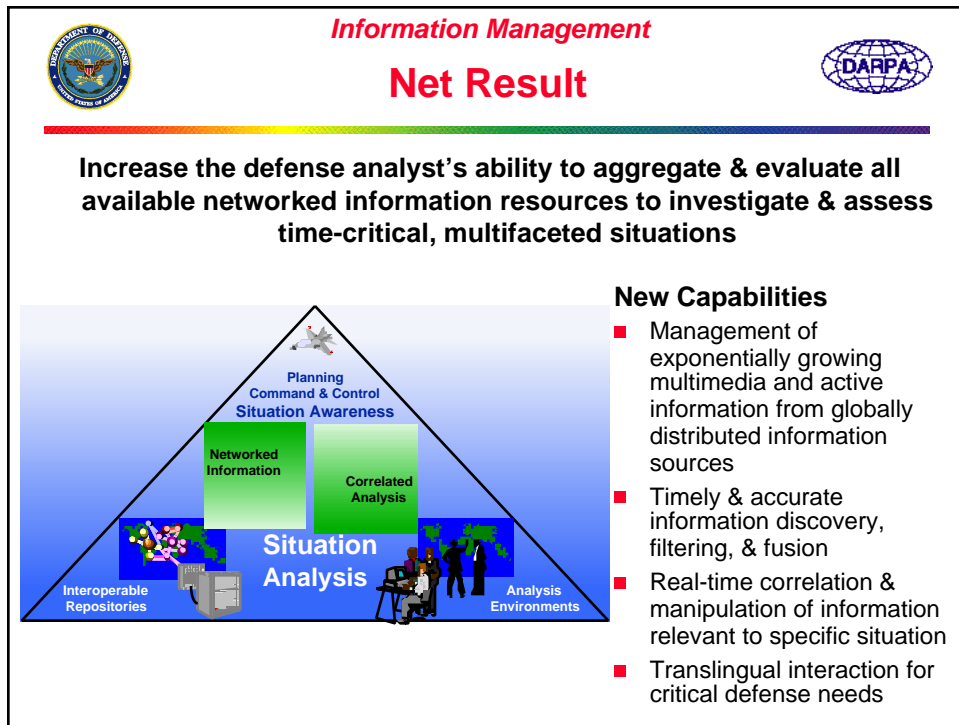
CNI - Coalition for Networked Information
DL - Digital Library
DTIC - Defense Technical Information Center
LC - Library of Congress

NARA - National Archives & Records Administration
NCSTRL - Networked Computer Science Technical Reports Library
PSOS - Prototype System of Systems
SUO - Small Unit Operations

The Integration Testbed is the principal transition point for technology developed under this program into Defense applications. This testbed activity will retain strong linkage into Information Systems Office (ISO) programs and to appropriate Service laboratories. But this is not the only means of technology transition relevant to Information Management. It is simply the point at which integrated test and evaluation can be conducted in Defense-related applications.

The work conducted on Analysis Environments will also result in broadly-available network-based services and freely-accessible software, continuing the tradition of maximizing return to society by broad dissemination of research results and products which has characterized DARPA's sponsored programs.

Designs and reference implementations for interoperable, secure repositories will, likewise be available to all. We anticipate that collaboration with the Library of Congress, the Defense Technical Information Center, and the National Archives and Records Administration will be very instrumental in ensuring that these designs are broadly applicable.



Perhaps one of the biggest mixed blessings confronting the defense analyst is the reality that information resources are growing exponentially in number and size.

The analyst's attention has become the critical resource. The objective of the Information Management program is to provide the technological capability to get the most out of the analyst's attention in the least amount of time.

The analyst's job, by definition, is to rapidly and effectively understand the full dimensions of an unfolding situation. Real-time correlation and manipulation of a broad array of information resources, as addressed in this program, is critical to this task.

Deriving a comprehensive and accurate assessment of an international situation draws heavily on the skills of translators and linguists. Translingual aids enable the analyst to perform substantial filtering of multilingual information, relaxing their reliance on translators while leveraging the precious skills of the translators to those tasks where their skills are more essential.

In conclusion, the Information Management program strives to broadly increase the defense analyst's ability to work with a diverse and distributed array of information resources in order to understand and develop an appropriate response to time critical, crisis-driven situations.